

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-196398

(43) 公開日 平成11年(1999) 7月21日

(51) Int.Cl.⁸

識別記号

F I

H 0 4 N 7/15

H 0 4 N 7/15

G 0 6 T 7/00

G 0 6 F 15/62

4 6 5 K

7/60

15/70

3 7 0

審査請求 未請求 請求項の数4 O L (全 10 頁)

(21) 出願番号

特願平9-359882

(22) 出願日

平成9年(1997)12月26日

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 一木 洋

東京都品川区北品川6丁目7番35号 ソニ

一株式会社内

(72) 発明者 近藤 哲二郎

東京都品川区北品川6丁目7番35号 ソニ

一株式会社内

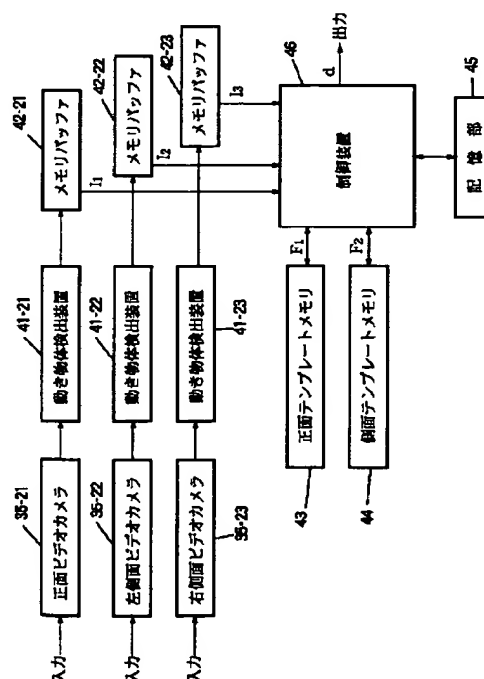
(74) 代理人 弁理士 稲本 義雄

(54) 【発明の名称】 画像処理装置

(57) 【要約】

【課題】 簡易に人の顔の向きを計測する。

【解決手段】 正面テンプレートメモリ43には、正面から撮影された顔(F1)が記憶されており、側面テンプレートメモリ44には、左側面から撮影された顔(F2)が記憶されている。各ビデオカメラは、それぞれ隣り合うビデオカメラと所定の角度(d I1乃至d I3とする)を有して設置されている。動き物体検出装置41-21乃至41-23は、それぞれ対応するビデオカメラから出力された画像から、人以外をマスキング処理し、対応するメモリバッファ42-21乃至42-23に出力する。制御装置46は、F1と各メモリバッファ42-21乃至42-23からの出力データI1乃至I3との相関度を、それぞれ計算する。そして、1番強い相関度のI1乃至I3と、F1との相関度R1、次に強い相関度のI1乃至I3と、F2との相関度R2を計算する。このR1、R2、およびd I1乃至d I3を用いて人の顔の向きを決定する。



【特許請求の範囲】

【請求項 1】 それぞれ隣り合うビデオカメラと所定の角度を有して設置されている複数のビデオカメラを用いる画像処理装置において、

被写体正面の画像を記憶する正面記憶手段と、
前記被写体側面の画像を記憶する側面記憶手段と、
前記複数のビデオカメラにより撮影された各画像のうち、前記正面記憶手段に記憶されている画像に 1 番近い画像と、2 番目に近い画像とを検索する検索手段と、
前記 1 番近い画像と、前記正面記憶手段に記憶されている画像との相関度を計算する第 1 の計算手段と、
前記 2 番目に近い画像と、前記側面記憶手段に記憶されている画像との相関度を計算する第 2 の計算手段と、
前記ビデオカメラの所定の角度、前記第 1 の計算手段により計算された値、および前記第 2 の計算手段により計算された値から、前記被写体の角度を決定する決定手段とを備えることを特徴とする画像処理装置。

【請求項 2】 前記正面記憶手段と前記側面記憶手段に記憶される画像は、前記複数のビデオカメラのうち、所定のビデオカメラに対して前記被写体が正面を向いている時に撮影された画像であることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 3】 前記側面記憶手段に記憶される画像は、前記複数のビデオカメラにより人の左側から撮影された顔、または、右側から撮影された顔であることを特徴とする請求項 2 に記載の画像処理装置。

【請求項 4】 前記側面記憶手段に記憶される画像は、前記複数のビデオカメラにより人の左側から撮影された顔と右側から撮影された顔との平均を取った顔の画像を記憶することを特徴とする請求項 2 に記載の画像処理装置。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】 本発明は画像処理装置に関し、特に、複数のビデオカメラを用いて人の顔の向きを計測するのに用いて好適な画像処理装置に関する。

【0002】

【従来の技術】 現在、複数の会議室をネットワークを用いて接続し、あたかも 1 つのテーブルを囲んで会議しているような遠隔会議システムがある。このようなシステムにおける各会議室は、例えば、図 7 のような構成になされている。この構成例は、4 つの会議室がネットワークにより接続されている場合を示している。ディスプレイ装置 21 乃至 23 は、会議室 4 にいる参加者 24 以外の参加者の映像と発言を出力するようになされている。例えば、ディスプレイ装置 21 のディスプレイ部 37-21 には、会議室 1 に存在する参加者の映像が表示されて、スピーカ部 38-21 には、その参加者の発言が出力される。また同様に、ディスプレイ装置 22 のディスプレイ部 37-22 には、会議室 2 に存在する参加者の

映像が表示され、スピーカ部 38-22 には、その参加者の発言が出力される。さらに、ディスプレイ装置 23 のディスプレイ部 37-23 には、会議室 3 に存在する参加者の映像が表示され、スピーカ部 38-23 には、その参加者の発言が出力される。

【0003】 ディスプレイ装置 21 乃至 23 上に設置されたビデオカメラ 51-21 乃至 51-23 は、参加者 24 の姿を撮影する。この際、各ビデオカメラに撮影された画像を、そのビデオカメラが設置されたディスプレイ装置に対応する会議室に送信するか（従って、それぞれの会議室に設置されているディスプレイ部には、それぞれ異なる参加者 24 の画像が表示される）、または、3 台のビデオカメラのうち、1 台のビデオカメラに撮影された画像を選択し、各会議室に送信するか（従って、それぞれの会議室に設置されているディスプレイ部には、同一の参加者 24 の画像が表示されている）、どちらかである。

【0004】 マイクホン 36-21 乃至 36-23 は、参加者 24 の発言を取り込み、各会議室の会議室 4 用のディスプレイ装置に設置されたスピーカ部に送信する。そして、そのスピーカ部からは、送信された参加者 24 の発言が出力される。

【0005】 指向性を有するマイクロホンを用いると、参加者 24 が、発言を伝えたい参加者の方を向いて発言すれば、その向かれた位置に設置されているマイクロホンが、他のマイクロホンに比べて、より大きなレベルで発言を取り込むことができる。従って、各会議室に設置されたスピーカ部から参加者 24 の発言が出力される際、その出力レベルは異なる。これにより、出力レベルの高いスピーカ部を有する会議室にいる参加者は、直感的に、参加者 24 は自分に向かって話をしていることを理解できる。

【0006】

【発明が解決しようとする課題】 上述した会議室の構成においては、複数のビデオカメラが設置されており、これらのビデオカメラを用い、参加者 24 の顔の向きを計測し、その計測した結果から、参加者 24 が、どの会議室にいる参加者に対して話しているのか判断し、その判断結果に対応した会議室にのみ、参加者 24 の発言を伝送したり、または、出力レベルを高くて伝送したりすることが可能である。

【0007】 しかしながら、このように参加者 24 の顔の向きを計測するには、立体計測などの複雑な処理を行う必要があるが、会議システムでは、厳密に顔の向きを測定する必要はなく、もっと簡単に顔の向きを計測することが望まれている。

【0008】 本発明はこのような状況に鑑みてなされたものであり、複数のビデオカメラを用いて、簡易に参加者の顔の向きを計測できるようにするものである。

【0009】

【課題を解決するための手段】請求項1に記載の画像処理装置は、被写体正面の画像を記憶する正面記憶手段と、被写体側面の画像を記憶する側面記憶手段と、複数のビデオカメラにより撮影された各画像のうち、正面記憶手段に記憶されている画像に1番近い画像と、2番目に近い画像とを検索する検索手段と、1番近い画像と、正面記憶手段に記憶されている画像との相関度を計算する第1の計算手段と、2番目に近い画像と、側面記憶手段に記憶されている画像との相関度を計算する第2の計算手段と、ビデオカメラの所定の角度、第1の計算手段により計算された値、および第2の計算手段により計算された値から、前記被写体の角度を決定する決定手段とを備えることを特徴とする。

【0010】請求項1に記載の画像処理装置においては、被写体正面の画像が記憶され、被写体側面の画像が記憶され、複数のビデオカメラにより撮影された各画像のうち、記憶されている被写体正面の画像に1番近い画像と、2番目に近い画像とが検索され、1番近い画像と、記憶されている正面画像との相関度が計算され、2番目に近い画像と、記憶されている側面画像との相関度が計算され、これらの計算された計算値と、ビデオカメラの所定の角度とから、被写体の角度が決定される。

【0011】

【発明の実施の形態】以下に本発明の実施の形態を説明するが、特許請求の範囲に記載の発明の各手段と以下の実施の形態との対応関係を明らかにするために、各手段の後の括弧内に、対応する実施の形態（但し一例）を付加して本発明の特徴を記述すると、次のようになる。但し勿論この記載は、各手段を記載したものに限定することを意味するものではない。なお、従来の場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【0012】請求項1に記載の画像処理装置は、被写体正面の画像を記憶する正面記憶手段（例えば、図4の正面テンプレートメモリ43）と、被写体側面の画像を記憶する側面記憶手段（例えば、図4の側面テンプレートメモリ44）と、複数のビデオカメラにより撮影された各画像のうち、正面記憶手段に記憶されている画像に1番近い画像と、2番目に近い画像とを検索する検索手段（例えば、図6のステップS5）と、1番近い画像と、正面記憶手段に記憶されている画像との相関度を計算する第1の計算手段（例えば、図6のステップS8）と、2番目に近い画像と、側面記憶手段に記憶されている画像との相関度を計算する第2の計算手段（例えば、図6のステップS8）と、ビデオカメラの所定の角度、第1の計算手段により計算された値、第2の計算手段により計算された値から、被写体の角度を決定する決定手段（例えば、図6のステップS9）とを備えることを特徴とする。

【0013】請求項2に記載の画像処理装置は、正面記

憶手段と側面記憶手段に記憶される画像は、複数のビデオカメラのうち、所定のビデオカメラ（例えば、図5の正面ビデオカメラ35-21）に対して被写体（例えば、図5の参加者24）が正面を向いている時に撮影された画像であることを特徴とする。

【0014】図1は、本発明の画像処理装置を適用したテレビ会議システムの構成を示している。なお、本明細書において、システムとは、複数の装置で構成される全体的な装置を表すものとする。図1に示されるように、複数（この実施例の場合、4つ）の通信センタ1-1乃至1-4が、例えば、ISDN（Integrated Service Digital Network）などのネットワーク2を介して相互に接続されている。また、各通信センタは、例えば、図2に示すような1つの会議室を備えている。

【0015】図2に示す会議室においては、1つのテーブル10、1つの椅子及び3台のディスプレイ装置が設けられている。例えば、通信センタ1-4の会議室においては、図2に示される番号4の位置に椅子が配置されており、番号1乃至3の位置にディスプレイ装置が配置されている。また、通信センタ1-3の会議室においては、図2に示される番号3の位置に椅子が配置されており、番号1、2、および4の位置にディスプレイ装置が配置されている。また、通信センタ1-2の会議室においては、図2に示される番号2の位置に椅子が配置されており、番号1、3、および4の位置にディスプレイ装置が配置されている。さらに、通信センタ1-1の会議室においては、図2に示される番号1の位置に椅子が配置されており、番号2乃至4の位置にディスプレイ装置が配置される。

【0016】また、通信センタ1-4の会議室に配置されたディスプレイ装置において、図2に示される番号1の位置に配置されたディスプレイ装置には、通信センタ1-1の参加者を撮影した画像が表示され、番号2の位置に配置されたディスプレイ装置には、通信センタ1-2の参加者を撮影した画像が表示され、番号3の位置に配置されたディスプレイ装置には、通信センタ1-3の参加者を撮影した画像が表示されるようになされている。他の通信センタも同様に、参加者が座る椅子以外の位置に配置されたディスプレイ装置には、対応する通信センタの参加者を撮影した映像が表示されることになる。

【0017】このように、それぞれの通信センタの会議室には、その通信センタの特定の位置に参加者が座るための椅子が配置されており、その他は、他の通信センタの参加者を表示するディスプレイ装置が配置されることになる。したがって、このように、会議室を構成することにより、どの通信センタの会議室においても、参加者の配置が同一の位置となる。すなわち、あたかも、テーブル10を中心にして、4人の参加者が、実際に特定の位置に配置されたような状態となる。ただし、各通信セ

ンタの会議室において、実在する参加者自身以外は、全てディスプレイ表示による参加者になるが、どの会議室においても、同様の会議状態が実現されていることになる。

【0018】次に、各通信センタの詳細について、図3を用いて説明する。なお、各通信センタとも、ディスプレイ装置の配置状態は多少異なるが、ほぼ同一の構成であるため、ここでは、通信センタ1-4についてののみ説明し、他の通信センタ1-1乃至1-3の説明は省略する。

【0019】まず、通信センタ1-4の会議室には、図2に示したように、番号4の位置に椅子が配置されており、番号1乃至3の位置にディスプレイ装置がそれぞれ配置されている。従って、図3に示される参加者24は、図2の番号4の位置に配置されている椅子に座ることになる。また、各ディスプレイ装置21乃至23には、他の通信センタからそれぞれ供給される画像を表示するディスプレイ部37-21乃至37-23、その画像に対応する音声を出力するスピーカ部38-21乃至38-23、およびその参加者24の発言を取り込むマイクロホン36-21乃至36-23が設けられている。さらに、参加者24の映像を取り込むために、参加者24の左側に設置されている左側面ビデオカメラ35-22、参加者24の正面に設置されている正面ビデオカメラ35-21、および参加者24の右側に設置されている右側面ビデオカメラ35-23が設けられている。

【0020】ディスプレイ部37-21乃至37-23とスピーカ部38-21乃至38-23は、通信センタ1-1乃至1-3から送信されてきた画像とその画像に対応する音声をそれぞれ出力するようになされている。すなわち、例えば、ディスプレイ装置21のディスプレイ部37-21には、通信センタ1-1の参加者の画像が表示され、スピーカ部38-21からは、その参加者の発言が出力されるようになされている。また、ディスプレイ装置22のディスプレイ部37-22には、通信センタ1-2の参加者の画像が表示され、スピーカ部38-22からは、その参加者の発言が出力されるようになされている。さらに、ディスプレイ装置23のディスプレイ部37-23には、通信センタ1-3の参加者の画像が表示され、スピーカ部38-23からは、その参加者の発言が出力されるようになされている。

【0021】また、ディスプレイ装置21に配置されている正面ビデオカメラ35-21は、通信センタ1-4の参加者24を撮影し、マイクロホン36-21は、その参加者24の発言を取り込み、その参加者24の画像と発言が、通信センタ1-1に供給される。また、ディスプレイ装置22に設置されている左側面ビデオカメラ35-22は、通信センタ1-4の参加者24を撮影し、マイクロホン36-22は、その参加者の発言を取

り込み、その参加者24の画像と発言が、通信センタ1-2に供給される。さらに、ディスプレイ装置23に設置されているビデオカメラ35-23は、通信センタ1-4の参加者24を撮影し、マイクロホン36-23は、その参加者の発言を取り込み、その参加者24の画像と発言が、通信センタ1-3に供給される。

【0022】そして、図3に示すように、ディスプレイ装置21乃至23は、参加者24が各ディスプレイ装置21乃至23のディスプレイ部37-21乃至37-23を見ることができるよう、図2に示された所定の位置に配置されている。

【0023】図4は、参加者24の顔の向きを計測するのに必要な装置を示したブロック図である。正面ビデオカメラ35-21に入力された画像は、動き物体検出装置41-21に出力される。動き物体検出装置41-21では、入力された画像から、動いている物体（この例では、参加者24）を検出し、その他の画像をマスキング処理し、メモリバッファ42-21に出力する。メモリバッファ42-21は、入力された画像を記憶する。同様に、左側面ビデオカメラ35-22に入力された画像は、動き物体検出装置41-22に出力され、人物以外の画像がマスキング処理され、メモリバッファ42-22に出力される。そして、メモリバッファ42-22に入力された画像は、一旦記憶される。さらに、右側面ビデオカメラ35-23に入力された画像は、動き物体検出装置41-23に出力され、人物以外の画像がマスキング処理され、メモリバッファ42-23に出力される。そして、メモリバッファ42-23に入力された画像は、一旦記憶される。

【0024】以上のようにメモリバッファ42-21乃至42-23に記憶された画像は、制御装置46の要請により、制御装置46に出力され、正面テンプレートメモリ43と側面テンプレートメモリ44に記憶されている各テンプレートを用いて、参加者24の顔の向きが測定され、出力される。記憶部45には、参加者24の顔の向きを計測する際に必要なデータが、適宜記憶される。

【0025】なお、側面テンプレートメモリ44は、左側面用と右側面用の2つのテンプレートメモリを用意しても良い。しかしながら、観測者24の顔の向きを計測する場合、人の顔がほぼ左右対称であることを利用すると、左側面用と右側面用の2つのテンプレートを用いる必要はない。従って、ここでは1つの側面テンプレートメモリ44を用いて顔の向きの計測を行う場合を説明する。この側面テンプレートメモリ44に記憶されるテンプレートは、左側面ビデオカメラ35-22または右側面ビデオカメラ35-23により撮影された画像である。また同様に、正面テンプレートメモリ43に記憶されるテンプレートは、正面ビデオカメラ35-21により撮影された画像である。

【0026】次に、正面ビデオカメラ35-21、左側面ビデオカメラ35-22、および右側面ビデオカメラ35-23の配置について、図5を参照して説明する。この例では、左側面ビデオカメラ35-22には、動き物体検出装置41-22とメモリバッファ42-22が内蔵されている。また、右側面ビデオカメラ35-23には、動き物体検出装置41-23とメモリバッファ42-23が内蔵されている。そして、正面ビデオカメラ35-21には、動き物体検出装置41-21、メモリバッファ42-21、正面テンプレートメモリ43、側面テンプレートメモリ44、記憶部45、および制御装置46が内蔵されている。また、各ビデオカメラは図示されていないケーブルなどにより、データのやりとりを行うことが可能になされている。

【0027】角度 dI_1 、 dI_2 、 dI_3 は、それぞれ正面ビデオカメラ35-21、左側面ビデオカメラ35-22、右側面ビデオカメラ35-23の角度を表している。ここでは、正面ビデオカメラ35-21の角度 dI_1 を0度とする。そして、左側面ビデオカメラ35-22と、正面ビデオカメラ35-21との角度 dI_2 をマイナス45度とし、ビデオカメラ35-23と、正面ビデオカメラ35-21との角度 dI_3 を45度とする。そして、参加者24の角度 dI_1 に対しての角度を角度 d (-45度乃至45度)とする。

【0028】次に、上述した3台のビデオカメラを用いて参加者24の顔の向きを計測する場合に付いて、図6のフロチャートを参照し、説明する。ここでは、側面テンプレートメモリ44に記憶されるテンプレートとしては、左側面ビデオカメラ35-22で撮影された画像とする。

【0029】ステップS1において、正面ビデオカメラ35-21の制御装置46は正面テンプレートメモリ43(F1)と側面テンプレートメモリ44(F2)の初期化を行うか否かを判断する。初期化は、システムの起動時に行われる。このシステムが起動される時は、4つの通信センタのそれぞれに参加する参加者の中の1人が、一番最初に通信センタのテレビ会議システムの会議室の照明を点灯させたとき、システムが起動される。つまり、例えば、通信センタ1-4の参加者24が、他の通信センタに比べて一番早く自分自身の通信センタ1-4の会議室の照明を点灯させたとき、その点灯に同期して通信センタ1-4の制御装置(不図示)が各通信センタに対して起動命令を送信する。そして、他の通信センタ1-1乃至1-3の制御装置が、この起動命令に対応して起動処理を行うことにより、各通信センタ内の装置の起動が完了される。

【0030】なお、上述の説明では、会議室の照明の点灯に同期して、起動命令を送信するようにしたが、例えば、最初の参加者が会議室のドアを開けたときに同期するようにしたり、最初の参加者が座席に座ったときに同

期するようにしたり、最初の参加者がシステムの電源を入れたときに同期するようにしてもよい。また、その他の方法によって、起動命令を送信してもよい。

【0031】ステップS1において、テンプレートメモリを初期化すると判断された場合、換言すれば、システムが起動された場合、ステップS4に進み、テンプレートメモリが初期化される。参加者24の正面に設置されているビデオカメラ35-21により撮影された画像は、正面テンプレートメモリ43に記憶される。また、左側面ビデオカメラ35-22により撮影された画像は、側面テンプレートメモリ44に記憶される。正面テンプレートメモリ43に記憶された画像をF1とし、側面テンプレートメモリ44に記憶された画像をF2とする。なお、ここで各テンプレートメモリに記憶される画像は、動き物体検出装置41-21(41-22)によりマスキング処理され、メモリバッファ42-21(42-22)に一旦記憶され、出力された画像である。

【0032】ステップS4において、テンプレートメモリの初期化が終了されると、ステップS2に進む。また、ステップS1において、テンプレートメモリの初期化を行わないと判断された場合も、ステップS2に進む。ステップS2においては、メモリバッファ42-22に記憶されている画像データI2とメモリバッファ42-23に記憶されている画像データI3との相関の度合い R_0 を計算する。相関の度合いの計算の仕方としては、パターンマッチングなどに用いられる最小2乗誤差を用いる。誤差が少なければ、I2とI3の相関が強いことを示しており、相関の度合い R_0 は大きな値を示すことになる。また、I2とI3の相関が強いということは、左側面ビデオカメラ35-22により撮影された画像と、右側面ビデオカメラ35-23により撮影された画像が、近似であることを示しているため、人間の顔がほぼ左右対称であることから、この場合、参加者24は、正面(正面ビデオカメラ35-21の方)を向いていることになる。

【0033】ステップS2において相関の度合い R_0 が計算し終わると、ステップS3に進み、しきい値 Th_0 と、相関の度合い R_0 の大小関係が判断される。このしきい値 Th_0 は、画像の解像度や階調により異なり、予め設定されている値とする。ユーザが任意の値を設定、変更するようにしても良い。このしきい値 Th_0 が、相関の度合い R_0 よりも小さい場合、ステップS4に進み、テンプレートメモリの更新が行われる。換言すると、参加者24が正面を向いている場合、相関の度合い R_0 が大きくなるので、しきい値 Th_0 よりも大きくなり、ステップS4に進みテンプレートメモリが更新される。このようにすることで、テンプレートメモリに記憶される画像は、参加者24が正面を向いている時の画像になる。

【0034】一方、ステップS3において、しきい値 T

h_0 が相関の度合い R_0 よりも大きいと判断された場合、ステップS5に進む。この場合、正面テンプレートメモリ43に記憶されている画像は、正面から撮影された参加者24の画像であり、側面テンプレートメモリ44に記憶されている画像は、左斜め45度の角度から撮影された参加者24の画像である。

【0035】ステップS5において、正面テンプレートメモリ43に記憶されているテンプレート F_1 と、メモリバッファ42-21に記憶されている画像 I_1 との相関の度合い、メモリバッファ42-21に記憶されている画像 I_2 との相関の度合い、およびメモリバッファ42-21に記憶されている画像 I_3 との相関の度合いを、それぞれ、ステップS2と同様に最小2乗誤差により求める。その求められた相関の度合いが強い順に N_1 、 N_2 、 N_3 とする。制御装置46は、このうち N_1 と N_2 を、記憶部45に記憶させる。この N_1 乃至 N_3 は、メモリバッファ42-21乃至42-23のいずれかを表す番号である。

【0036】ステップS5において、 N_1 と N_2 が記憶部45に記憶されると、ステップS6に進む。ステップS6において、 $N_1=I_2$ かつ $N_2=I_3$ 、または、 $N_1=I_3$ かつ $N_2=I_2$ 、であるか否かが判断される。 $N_1=I_2$ かつ $N_2=I_3$ であるということは、参加者24が正面を向いている時の画像（正面テンプレートメモリ43に記憶されているテンプレート）と1番近い画像が、 I_2 （ I_2 は、参加者24の左側に設置されている左側面ビデオカメラ35-22により撮影された画像である）であり、かつ、次に参加者24が正面を向いている時の画像と近い画像であるのが、 I_3 （ I_3 は、参加者24の右側に設置されている右側面ビデオカメラ35-23により撮影された画像データである）であるということを示している。

【0037】正面テンプレート F_1 と相関の度合いが1番強いのと次に強いのは、必ず隣り合うビデオカメラにより撮影された画像である。従って、上述したような状態は存在しないので、正常な状態でないと判断され、ステップS7に進む。 $N_1=I_3$ かつ $N_2=I_2$ である場合も、同様の理由が成り立ち、これらの関係が成り立つと判断された場合、ステップS7に進む。

$$d = dN_1 + (dN_2 - dN_1) \times R_2 / (R_1 + R_2) \cdots (1)$$

ここで、 dN_1 は、 N_1 と判断されたメモリバッファが内蔵されているビデオカメラの角度 dI_1 乃至 dI_3 である。また、同様に dN_2 は、 N_2 と判断されたメモリバッファが内蔵されているビデオカメラの角度 dI_1 乃至 dI_3 である。

【0043】図5の場合を例にあげて説明すると、この例では、参加者24は、右側面ビデオカメラ35-23の方に、正面ビデオカメラ35-21（角度 dI_1 ）に対して角度 d だけ向いている。正面テンプレートメモリ

$$d = 45 + (0 - 45) \times R_2 / (R_1 + R_2)$$

【0038】ステップS7において、異常な状態と判断されたのは、3回目か否かが判断される。3回目ではないと判断された場合、ステップS5に戻り、相関の順番が計算し直される。換言すれば、ステップS6において、異常な状態であると判断されるのは、ステップS5における計算に間違いがあったと判断し、もう1度計算し直す。

【0039】一方、ステップS7において、異常な状態であると判断されたのは3回目であると判断された場合、ステップS4に進みテンプレートの更新が行われる。換言すれば、ステップS6において、異常な状態であると判断されるのは、正面テンプレートメモリ43に記憶されていたテンプレート（ F_1 ）がおかしいと判断し、もう1度テンプレートの更新を行う。

【0040】なお、上述した説明では、異常な状態と判断されたのは3回目か否かを判断したが、3回目でなく他の回数に設定しても良い。

【0041】一方ステップS6において、正常な状態であると判断された場合、ステップS8に進む。ステップS8において、ステップS2と同様に、最小2乗誤差を用いて、正面テンプレートメモリ43に記憶されている正面テンプレート F_1 と N_1 との相関の度合い R_1 、側面テンプレートメモリ44に記憶されている側面テンプレート F_2 と N_2 との相関の度合い R_2 、がそれぞれ計算される。なお、 N_1 と N_2 は、上述したように、メモリバッファ42-21乃至42-23を示す番号であり、実際に計算に用いられるのは、これらのメモリバッファに記憶されている I_1 乃至 I_3 のデータである。従って、 F_1 と N_1 の相関の度合い R_1 とは、換言すれば、 F_1 と N_1 が示す I_1 乃至 I_3 のいずれかデータとの相関の度合い R_1 ということになる。よって、相関の度合い R_1 は、ステップS5において既に計算されている値を用いることができ、その値を記憶部41に記憶させておくことにより、ステップS7において、相関の度合い R_1 を計算する必要はなくなる。

【0042】ステップS7において、相関の度合い R_1 、 R_2 が計算し終わると、ステップS8に進み、参加者24の顔の角度 d が次式に従って、計算される。

43に記憶されている正面テンプレート F_1 と最も相関の強い画像を出力するのは、右側面ビデオカメラ35-23である。従って、この場合、 $dN_1=dI_3=45$ 度になる。また、次に正面テンプレートメモリ36に記憶されているテンプレート F_1 と相関の強い画像を出力するのは、正面ビデオカメラ35-21である。従って、この場合、 $dN_2=dI_1=0$ 度になる。

【0044】上述した値を式(1)に代入すると、以下のようになる。

$$= 45 + (-45) \times R_2 / (R_1 + R_2) \cdots (2)$$

となる。

【0045】このように求められた参加者の顔の向きを利用し、例えば、特定の参加者に対して発言を行う場合、その参加者が表示されているディスプレイ装置の方に向かって発言することにより、その参加者のみにその発言を伝達させることも可能となる。

【0046】上述した説明においては、人の顔が左右対称であることを利用し、1つの側面テンプレートメモリ44を用いて参加者24の顔の向きの計測を行ったが、左側面用と右側面用の、それぞれのテンプレートメモリを用意し、これらを用いて上述した顔の計測を行うことも可能である。また、側面テンプレートメモリ44に記憶させる画像として、左側面ビデオカメラ35-22と右側面ビデオカメラ35-23により撮影された顔の画像の平均を取り、その平均を取られた画像を記憶するようにしても良い。

【0047】上述した説明では、4つの通信センタのすべてに参加者がいる状態でのテレビ会議の実行方法について説明したが、例えば、通信センタ1-1に参加者がいない場合、その参加者がいない状態での画像が各通信センタの対応するディスプレイ装置のディスプレイ部に表示されることになる。

【0048】なお、上述した説明において、通信センタが4つである場合について説明したが、本発明はこれに限らず、通信センタをさらに多くしてもよいし、少なくしてもよい。また、本発明の主旨を逸脱しない範囲において、さまざまな変形や応用例が考えうる。従って、本発明の要旨は、実施例に限定されるものではない。

【0049】

【発明の効果】請求項1に記載の画像処理装置によれば、被写体正面の画像を記憶し、被写体側面の画像を記憶し、複数のビデオカメラにより撮影された各画像のうち、記憶されている被写体正面の画像に1番近い画像

と、2番目に近い画像とを検索し、1番近い画像と、記憶されている正面画像との相関度を計算し、2番目に近い画像と、記憶されている側面画像との相関度を計算し、これら計算された計算値と、ビデオカメラの所定の角度から、被写体の角度を決定するようにしたので、簡易な被写体の向きの計測が可能となる。

【図面の簡単な説明】

【図1】本発明の画像処理装置を適用したテレビ会議システムの構成例を示す図である。

【図2】本発明のテレビ会議システムにおける各通信センタの会議室の状態を示す図である。

【図3】図1における通信センタにおけるディスプレイ装置の配置状態を示す図である。

【図4】図3における各ビデオカメラで撮影された画像の流れを説明する図である。

【図5】図3におけるビデオカメラの配置状態を示す図である。

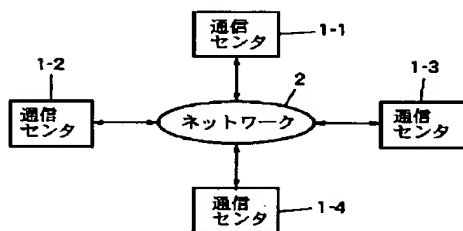
【図6】参加者の顔の向きの計測を説明するフローチャートである。

【図7】従来のテレビ会議システムにおける各通信センタにおけるディスプレイ装置の一例の配置状態を示す図である。

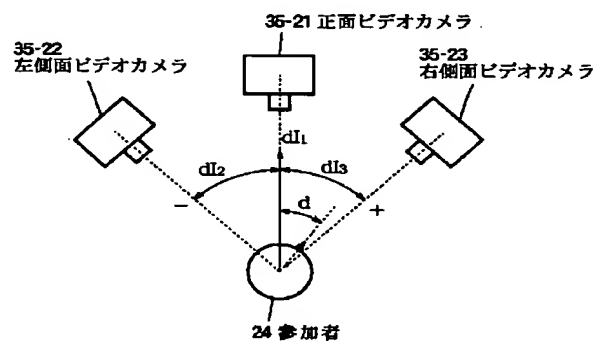
【符号の説明】

1-1乃至1-6 通信センタ, 2 ネットワーク,
21-23 ディスプレイ装置, 35-21 正面
ビデオカメラ, 35-22 左側面ビデオカメラ,
35-23 右側面ビデオカメラ, 36-21乃至3
6-23 マイクロホン, 37-21乃至37-23
ディスプレイ部, 38-21乃至38-23 スピー
カ部, 41-21乃至41-23 動き物体検出装
置, 42-21乃至42-23 メモリバッファ,
43 正面テンプレートメモリ, 44 側面テンプレ
ートメモリ, 45 記憶部, 46 制御装置

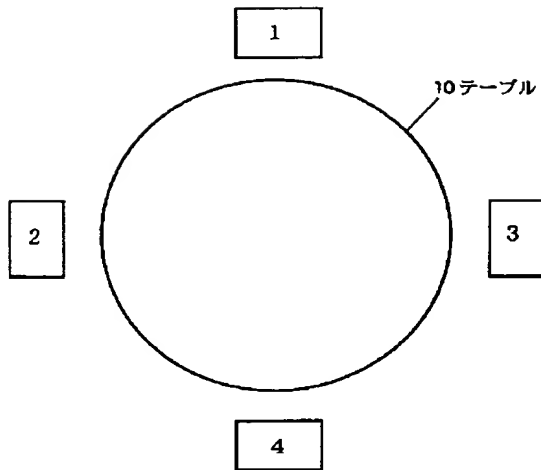
【図1】



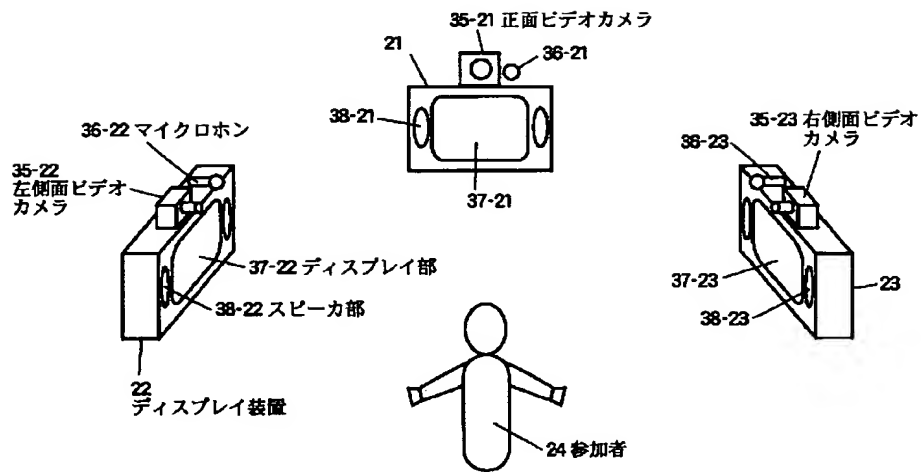
【図5】



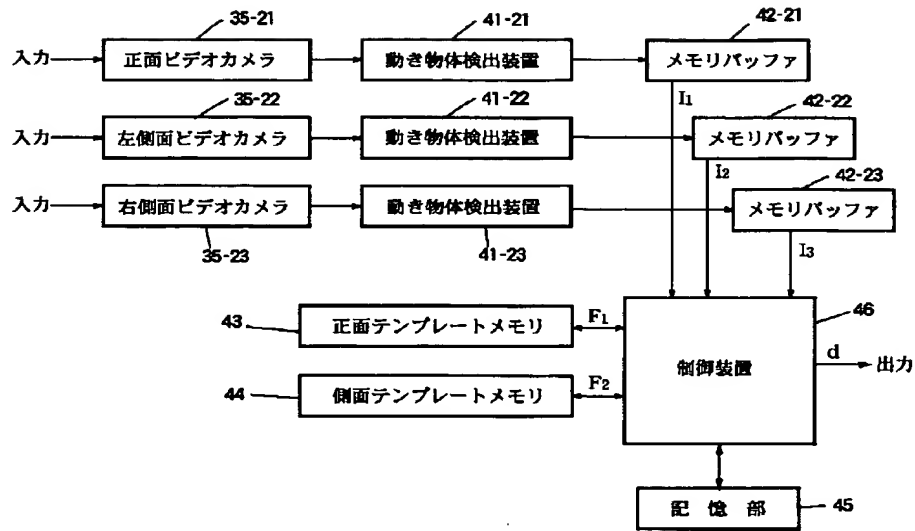
【図 2】



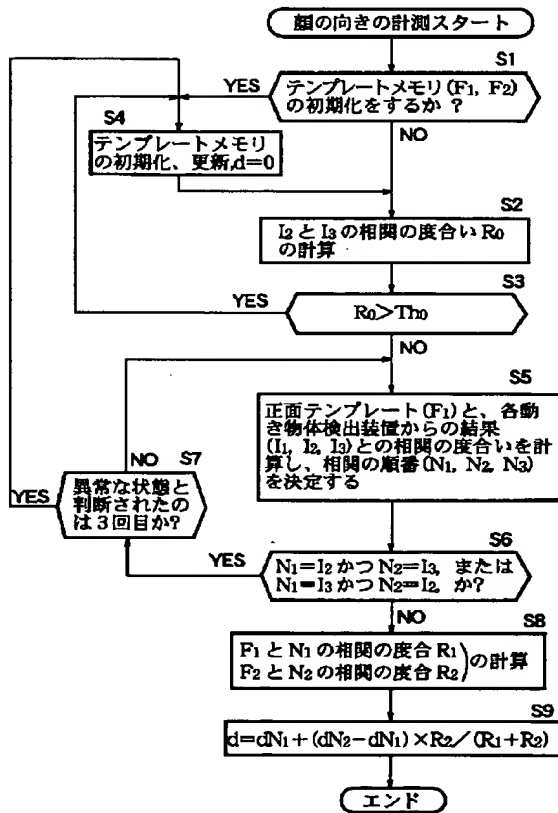
【図 3】



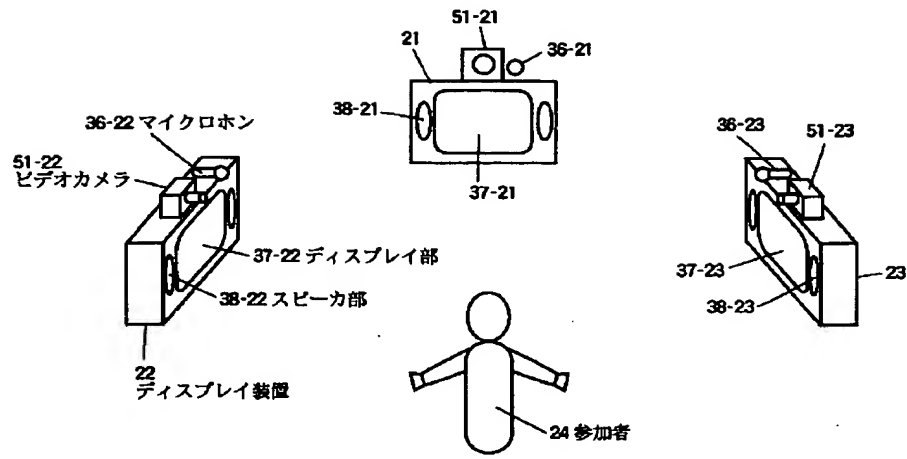
【図 4】



【図 6】



【図 7】



動画像を用いた顔認識システム

山口 修 福井 和広 前田 賢一

(株) 東芝 関西研究所

〒 658 神戸市東灘区本山南町 8-6-26

E-mail: osamu@krl.toshiba.co.jp

あ ら ま し 本稿では、動画像を利用した顔認識システムについて報告する。本研究では、登録から認識までの時間経過が短い場合を対象として、顔向きや表情変化といった変動を吸収する顔認識の方法について考察する。変動を吸収するために、単一の画像だけで認識するのではなく、時系列画像を用いた認識を行なう。入力画像列をK-L展開によって部分空間として表現し、登録パターンの部分空間との2つの部分空間の間の角度を類似度とする「相互部分空間法」を適用した。相互部分空間法の有効性を調べるために、画像中から高速に正規化顔画像パターンを切り出すシステムを用いて、大量の実験データを収集し、認識実験を行った。その結果について報告する。

キーワード 顔画像認識、動画像、相互部分空間法

Face Recognition System using Temporal Image Sequence

Osamu Yamaguchi, Kazuhiro Fukui and Ken-ichi Maeda

TOSHIBA Kansai Research Laboratories

8-6-26 Motoyama-Minami-Cho, Higashinada-ku,

Kobe 658 Japan

E-mail: osamu@krl.toshiba.co.jp

Abstract This paper presents a face recognition system for human interface and security using temporal image sequence. As input we utilize several images of a face rather than a single shot so that the influence of facial expression and face direction can be attenuated. For the identification of the face, we essentially compute the subspace from the image sequence and apply the "Mutual Subspace Method" that defines the similarity by the angle between the computed subspace and those of pre-registered faces. It is naturally assumed that a registered face retains its form. We generate numerous normalized face patterns by the system and demonstrate the effectivity through the experimental results.

key words Face recognition, Temporal image sequence, Mutual subspace method

1 はじめに

近年、ヒューマンインタフェース、セキュリティなどへの応用を目指して、顔を用いた個人認証技術が重要な要素技術として認知されている。顔画像認識を用いる利点として、非接触で、利用者に意識させずに認識可能といったことが挙げられる。

図1に示すように、システムの利用者が、作業を中断してシステムから少し離れるような状況を想定した場合、それまでシステムを利用していた人が、誰であるかを瞬時に判定することができれば、ヒューマンインタフェースや簡易なセキュリティ技術として利用可能である。この時、登録作業を明示的に指示して行なうのではなく、図1(a)のように、利用者が作業中の状態の時に顔画像を収集し、図1(b)の作業を中断して離れている間に登録処理を行ない、利用者が戻ってきた場合(図1(c))に認識を行なうことができれば、利用者への負担が少なく、より自然な状況での顔認識が可能になる。

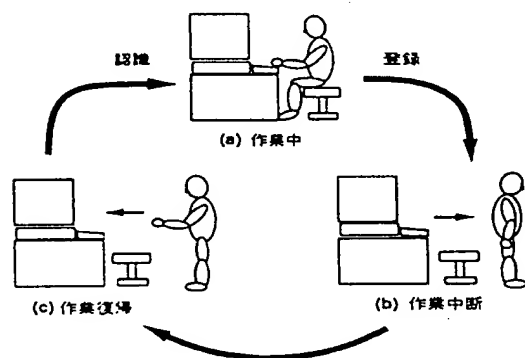


図1: 登録と認識の流れ

このようなシステムを実現するためには、(1)環境の変動は少ないものの、作業中の利用者の顔の向き変化、表情変化などがあり、これらの変動に影響を受けない安定した認識法、(2)高速で自動的に顔画像パターンを収集し、登録作業、認識処理を行なうための画像処理方法、が必要である。

本研究では、(1)については、動画画像から短時間に収集した複数枚の画像を用いて、顔の向き、表情などの変動を吸収する顔認識法について考察する。(2)に関しては形状情報とパターン情報を組合せにより、顔部品を安定に抽出する方法[1]を用いて、正規化顔画像パターンを高速に生成する。

以下本稿では、時系列画像を用いた相互部分空間法[2]による顔認識方法について述べる。次に、短時間で大量の多様な顔画像データを収集するために試作したシステムの概要と、各処理内容について説明する。最後に、図

1に示した、登録から認識までの時間経過が短い場合を想定して、実験データを収集し、評価実験を行った結果について報告する。

2 時系列画像を用いた顔認識

本稿では、表情変化や顔の向きといった変動に対応するため、動画画像から得られた時系列画像を用いた認識法について考える。動画画像を用いた顔認識の研究としては、特徴抽出のために動画画像の情報をを用いるものとして、まばたきを検出し、目の位置を求め、パターンを正規化して顔認識を行なう研究[3]がある。これは、個人識別のために積極的に動画画像情報を利用したものではない。また、顔の動き情報を抽出し、時間方向にコード化したものをIDとして認識を行なう研究[4]のように、時間的な構造情報を抽出して、識別に利用するものなどがある。

静止画を対象とした研究としては、表情変化や顔の向きといった変動を吸収することを目的として部分空間法を用いた研究[5]がある。各個人のパターン情報を登録する際に、様々な表情や顔の向きのバリエーションを考慮してデータを収集しておけば、部分空間法により、それらのデータをK-L展開(主成分分析)し、ベクトル空間として効率よく個人情報を記述できる。

本研究では、登録時だけでなく認識の際にも多数枚の多様なデータ(時系列画像)を収集して、認識に利用する。そのために、部分空間法を拡張した方法である相互部分空間法[2]を用いる。この方法は、登録パターンだけでなく、入力パターンも部分空間として表現し、登録パターンの部分空間と、2つの部分空間の間の角度を類似度として識別を行なう。

時系列画像を用いる利点をまとめると、1)顔に特有である表情変化、顔の向きなどの変化に対応、2)撮影時の画像ノイズなどの影響の減少、3)顔領域抽出の失敗など不完全なデータの混入の軽減、などが挙げられる。

1)については、顔の向きや表情変化は時間的に変動するものであり、ある時点での一枚だけを認識するのでは、不十分である。一定時間に数枚の画像取得を行うことで、入力パターンの冗長化をはかることで、上述したような変動要素を吸収することが期待できる。2)については、一般に同じカメラパラメータで撮影された n 枚の画像を平均した場合、その雑音レベルは $1/\sqrt{n}$ となることが知られており、単純に多数枚の画像を平均しただけでも、効果がある。3)については、画像処理の失敗によって、認識が正確に行われず排除されたり、誤った認識結果を求めることを軽減することができる。

2.1 相互部分空間法

部分空間法[6]では、類似度 $S_{normal}(g)$ は、 g を入力画像、 o を登録パターンを主成分分析して得られた各固

有ベクトル（辞書と呼ぶ）として、以下の式で与える。

$$S_{normal}(g) = \frac{1}{\|g\|^2} \sum_{n=0}^N (g \cdot \phi_n)^2 \quad (1)$$

(\cdot) は内積を表す。

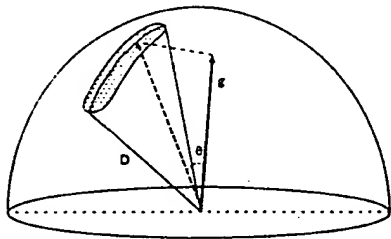


図 2: 部分空間法の説明図

図 2 は部分空間法の説明図であり、未知入力ベクトル g を部分空間 D に射影したベクトルとの角度、すなわち部分空間との角度から式 (1) のように類似度を定義している。

一方、相互部分空間法では、入力ベクトルも部分空間で表現し、辞書パターンの部分空間との間の角度を類似度として識別を行なう。図 3 は相互部分空間法の説明図であり、2 つの部分空間 D, G のなす角度 θ の余弦は、

$$\cos^2 \theta = \sup_{d \in D, g \in G, \|d\| \neq 0, \|g\| \neq 0} \frac{|(d, g)|^2}{\|d\|^2 \|g\|^2} \quad (2)$$

と定義する (d, g は式 (2) が極値をもつためのそれぞれの部分空間上のベクトルを表す。) これに関して、2 つの部分空間 D, G への正射影作用素を P, Q とする場合、 $\cos^2 \theta$ は PQP の最大固有値 λ_{max} となる [2]。

$$\cos^2 \theta = \lambda_{max} \quad (3)$$

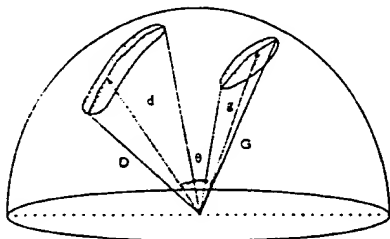


図 3: 相互部分空間法の説明図

辞書パターン部分空間を D 、入力された時系列画像に対する部分空間を G とする。ここで、 θ を各部分空間

D, G における固有ベクトルとする。実際には PQP という行列の最大固有値を求めるのではなく、式 (4) で表される行列 X の固有値問題を解き、その最大固有値を類似度（部分空間間類似度） S_{mutual} とすればよい [2]。ここで、 D の部分空間の次元を M 、 G の部分空間の次元を L として、 $L \leq M (1 \leq i, j \leq L)$ とする。

$$X = (x_{ij}) = \sum_{m=1}^M (\psi_i, \phi_m)(\phi_m, \psi_j) \quad (4)$$

$$W^T X W = \Lambda \quad (5)$$

($W: X$ の対角化行列、 $\lambda_{max}: \Lambda$ の対角成分の最大値)

$$S_{mutual}(G, D) = \lambda_{max} \quad (6)$$

本方法は、文字認識の高度化のために開発された方法であるが、メディアに依存しない方法であり、顔画像処理のように連続してデータが随時取得できるという点で、入力ベクトルの冗長化という点で親和性がよく、さまざまな変動にロバストな認識が可能になると考えられる。

3 顔認識システム概要

次に試作したシステムについて述べる。顔認識システムは、図 4 に示すように、(a) カメラから入力された画像中の顔領域の検出、(b) 顔部品（瞳、鼻孔）の特徴点抽出、(c) 位置、大きさなどを正規化したパターンの切りだし、(d) 登録パターン辞書との類似度計算による識別処理、を順に行なってシステムの前の人が誰であるかを識別する。

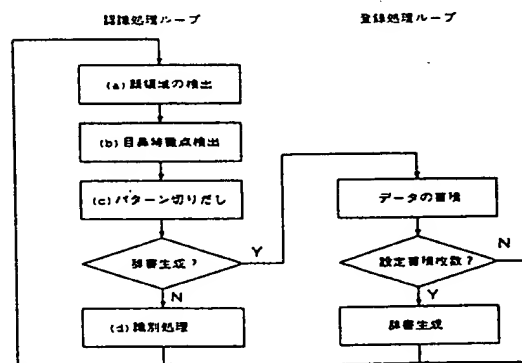


図 4: 処理の流れ図

試作したシステムでは、用意されたコントロールパネルウインドウを操作することにより、辞書生成を指示するだけで、認識処理から登録処理に移行でき、自動的に顔画像パターンの収集を行なう。辞書登録では、数枚の

パターンを蓄積し、個人を識別/同定するための辞書生成が行なわれる。登録処理が終了すると、図4のように再び認識処理を実行する。

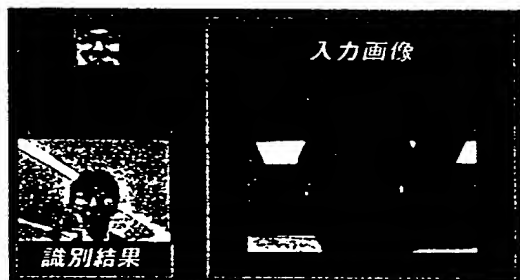


図5: 顔認識システムの概観(上: 画面表示、下: カメラの配置)

システムの実装はSGIワークステーションIndy(CPU R5000 150MHz)上で行なっており、システムはソフトウェアのみで実現されている。画像入力のためのカメラは、図5下のように、ディスプレイの下側に設置して、顔を撮影する。カメラは、東芝製IK-C40 CCDカラーカメラ、レンズは焦点距離9.5mmのものを使用した。取得画像のサイズは、320×240 pixelであり、カラーカメラを用いているが、特徴点抽出、認識には濃淡情報のみを利用している。

図5上は、システムの表示内容であり、入力画像の表示と、正規化パターンの表示、識別結果の表示を行なう。入力画像の表示部分には、顔検出、目鼻検出の様子を示している。認識が行なわれると、識別結果として登録時に撮影した顔画像を表示する。試作システムでは、部分空間法による識別を行っており、100人の識別を一秒間に3～4回で処理できる。なお、相互部分空間法による認証は実装中である。

次章で、システムによる正規化顔画像パターン的高速生成方法について説明する。

4 正規化顔画像パターンの高速生成

4.1 顔領域の検出

本システムでは、画像中に一人の人物のみが存在することを仮定する。顔領域の検出は、部分空間法[6]を用い

た、パターンマッチングを行う。予めさまざまな人物、方向の顔パターンデータ(15×15 pixelの画像)を収集し、顔検出用の辞書を作成しておく。そのテンプレートを画像全体にラスタースキャンし、各場所での類似度を逐次求め、類似度の局所最大値が、しきい値以上の箇所を顔領域として検出する。顔領域の大きさの変化を考慮して、入力画像を3段階にスケール変化させて、パターンマッチングを行い、画像中に一つの顔領域を決定する。画像中にしきい値以上の箇所が無い場合は、人が存在しないと判断し、以後の認識処理を行わない。

また、[8]のように、複雑な背景の中から顔を見つけるために、顔の辞書との類似度の高い非顔データを収集しておき、それらから構成された非顔辞書との類似度を同時に求めることによって、顔検出の誤認識を防いでいる。

顔領域を検出するために、動画像の利点を活かして、フレーム間差分の情報を用いて、移動領域のみを検出対象領域とする。これにより全画面探索を行なう必要がなく、高速化を計っている。

4.2 顔部品特徴点抽出

次に、検出された顔領域に対して、瞳、鼻孔の特徴点の抽出を行なう。従来法としては、Eigenfeature法[9]やカラー情報を用いた特徴点抽出法[10]などがある。いずれもパターン情報のみや領域形状情報のみを利用した方法であるが、本研究では、形状情報とパターン情報を組み合わせることによって、高速かつ、位置ずれの少ない特徴点抽出法[1]を用いる。まず、形状情報を用いた特徴点抽出には、明度変化、ノイズにロバストである[7]で提案された領域ベースのエッジ抽出法を拡張した、「分離度フィルタ」による抽出法を用いる。

分離度 S ($0.0 \leq S \leq 1.0$) は次の式で与えられる[7]。

$$S = \frac{n_1(\bar{P}_1 - \bar{P}_m)^2 + n_2(\bar{P}_2 - \bar{P}_m)^2}{\sum_{i=1}^N (P_i - \bar{P}_m)^2}$$

ここで、 N は領域内の全画素数、 n_1, n_2 は領域1、領域2の画素数、 P_k は位置 k における輝度、 \bar{P}_m は領域全体での輝度の平均値、 \bar{P}_1, \bar{P}_2 は領域1、領域2での輝度の平均値を示す。

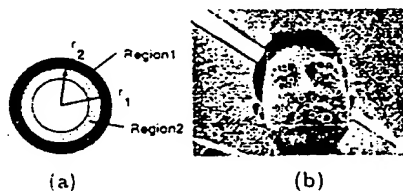


図6: 円形分離度マスク(a)と分離度の局所最大点(b)

図6(a)に示すような円形のマスクを設定し、そのマスク内で2つの円形領域の分離度を上述した式によって計算する。位置、半径を変えながら、円形の領域を特徴点候補として算出する。これを分離度フィルタと呼ぶ。先に求めた顔領域中で、分離度の局所最大点を候補として抽出した結果を図6(b)に示す。

これらの目鼻の特徴点候補を求めたのち、さらにパターン照合による検証を行って、正確な目鼻の位置を特定する。瞳、鼻孔に対しては額の検出と同様、あらかじめ収集した濃淡パターンから部分空間法のための辞書を構成しておく。そして、各特徴点において、目鼻辞書との類似度を求め、幾何学的配置の制約を満たす4点を特徴点として検出する。特徴点抽出についての詳細は、[1]に示す。

4.3 正規化パターンの切り出し

抽出された目鼻の特徴点を基準にして顔領域を位置、サイズを正規化し、矩形的濃淡パターンとして切り出す。

図7(a)の顔領域について、図7(b)のように両目の瞳を結ぶベクトル V_1 と、両目の瞳位置の midpoint から、鼻孔の midpoint を結ぶベクトル V_2 を、それぞれを $1/m_1, 1/m_2$ (m_1, m_2 は定数)した E_{V_1}, E_{V_2} を求める。図7(b)のCを基準点として、 E_{V_1}, E_{V_2} の線形和 $aE_{V_1} + bE_{V_2}$ (a, b は整数)した位置の画素を抽出し、図7(c)のように、 30×30 pixelの部分画像を生成する。

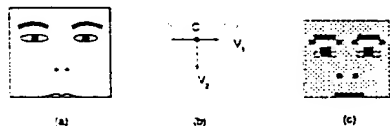


図7: 正規化パターンの生成

m_1, m_2 を変化させることにより、切り出しを行なう範囲を設定できる。これについては、目付近には、個人性をあらわす情報が多く含まれていること、髪型の変化などの影響を軽減すること、口の開閉の可能性がありパターンの変動が大きいこと、を考慮し、図7(a)に示されている範囲となるように設定した。

図8は、実際に連続して得られた20枚の矩形的濃淡パターンである。処理速度は一秒間に4~5枚の切り出しが可能で、自動的に大量のパターンが生成できる。

5 評価実験

日本人男女101名をワークステーションの前に座わらせて、データ収集を2回行なった(照明条件は一定)。1回目は、無表情で画面中を移動する指標に顔の方向を向けてもらい、180枚のデータを撮影した。その後、図



図8: 正規化パターン取得例

1のように、一度離席し、再度着席してもらい、2回目の画像取得を行なった。2回目の画像取得では、会話中(口の開閉)の画像や表情変化した画像などもデータとして含まれている。

正規化顔画像パターンは試作システムを用いてリアルタイムで収集し、識別評価実験についてはオフラインで行なった。辞書は1回目に取得した180枚を利用した部分空間とする。評価用の入力画像系列として2回目に取得した180枚を使用する。

5.1 識別実験

入力パターンとして得られた 30×30 pixelの画像を、濃淡の補正を行なうため、ヒストグラム平坦化処理を行なう。その画像をさらに計算量を削減するために、 30×30 から 15×15 に解像度を落とし、225次元のベクトルとして認識を行なう。

識別は、登録されている人の辞書(固有ベクトル集合)との類似度を、部分空間法によってすべて求め、最も高い類似度をとった辞書に対応した人を識別結果とする。識別正解率は、最大類似度をとった人と実際の入力データが同じ人である場合を正解として、正解数/入力データ数として定義する。

試作システムにおける、識別正解率は、類似度計算に利用する固有ベクトルの数 N を変化させた場合、表1のように推移する。

表1: 識別正解率と次元数の関係

| 次元数 | 5 | 10 | 15 |
|-------|-------|-------|-------|
| 識別正解率 | 89.9% | 92.1% | 92.9% |
| 次元数 | 20 | 25 | 30 |
| 識別正解率 | 93.3% | 93.5% | 93.7% |

5.2 同定実験

セキュリティに応用することを考えた場合、登録されていない他人を間違えて受け取ってしまったら、本人であるのに排除されてしまうということが起こってはならない。これらの評価は識別正解率ではなく、本人排除率(FRR: False Reject Error)/他人受取率(FAR: False Accept Error)で評価する。

以下の4つの実験を行った。

- 部分空間法による同定実験
- 相互部分空間法による同定実験
- 部分空間法+頻度による同定実験との比較
- 平均ベクトルを用いた部分空間法との比較

図9は、通常の部分空間法により、本人排除率/他人受取率を示したものである。本人であるかどうかは、登録辞書との類似度がしきい値を越えたか否かで判断する。横軸は、その設定しきい値、縦軸が、それぞれの率をパーセントを示している。試行回数は、本人 $101 \times 180 = 18180$ 回 / 他人 $100 \times 101 \times 180 = 1818000$ 回である。設定しきい値を高くすることで、本人排除率が上昇し、他人受取率が低下することがわかる。

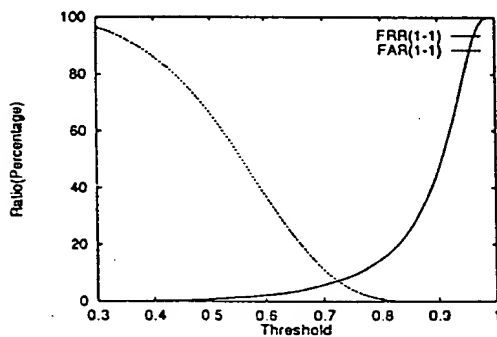


図9: 部分空間法による認識結果

次に、相互部分空間法の評価結果を示す。評価用画像の連続した30枚を用いて、入力部分空間を計算し、式(6)の定義のように類似度を求める。部分空間間類似度を求めるためのそれぞれの次元数は入力部分空間5次元/登録辞書部分空間5次元とした。それぞれ180枚のデータ取得を行なったため、30枚ずつ6セットのデータ系列を作成して評価を行なった。試行回数は、本人606回 / 他人60600回である。図10は、相互部分空間法による、本人排除率/他人受取率を示したものである。

図9と比較して、FRRとFARの交差点が降下しており、多数枚を用いた認識により性能が向上しているのがわかる。FRR (本人排除率)を0%とする点で比較すると、部分空間法がFAR: 約80%に対し、相互部分空間法は、FAR: 約30%であり、他人排除にも効果があることが示された。

図11は、ある一人の実験データについて、各方法の類似度の時系列変化を比較したものである。相互部分空間法については、 i フレームから $i+29$ フレームまでのデータから入力部分空間を構成し、計算を行なった結果

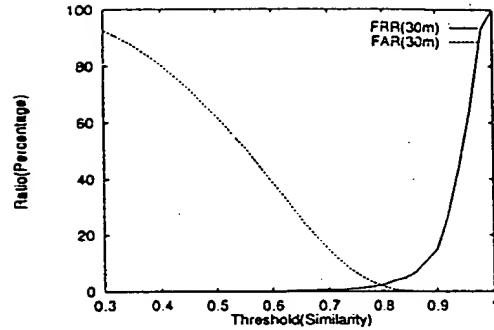


図10: 相互部分空間法による認識結果

を示している ($0 \leq i \leq 150$)。部分空間法によって1枚ずつ計算される類似度が、各フレーム毎に変化が大きいものに対して、相互部分空間法では、大きな変化を示さず、入力データの冗長化によって変動を吸収しているのが分かる。

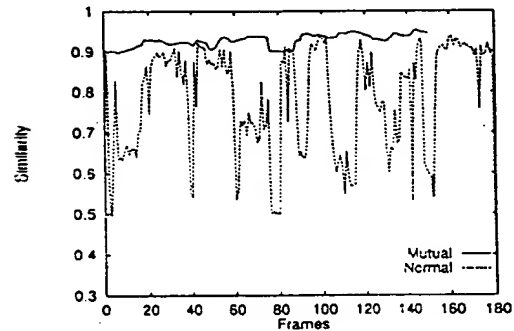


図11: 類似度の時系列変化の比較 (実線: 相互部分空間法、破線: 部分空間法)

次に比較実験として、通常の部分空間法を30回連続行ない、本人と同定される頻度を用いて認識を行なった結果を示す。ここでは、30回の認識のうち、10回以上類似度が閾値を越えた場合に本人と同定する。この10回の決め方は、FAR/FRRの交差点が相互部分空間法と同程度になるように設定した。結果を図12に示す。

図10の相互部分空間法と比較すると、わずかであるがFARが大きくなっており、入力側の部分空間を生成する際に行なわれるK-L展開によって、データ間の相関が考慮されるために、認識率の向上に寄与していると考えられる。

最後に時系列画像の平均ベクトルを用いた部分空間法と相互部分空間法の比較を行なう。試行回数は、6つに

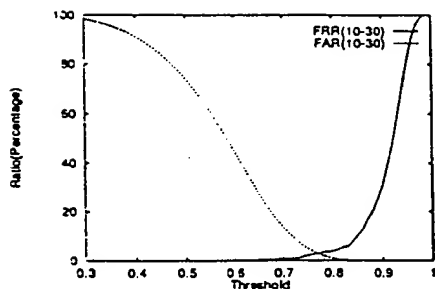


図 12: 部分空間法+頻度を用いた結果

分割した1セットのみを使用したため、先の結果に比べて少なく 本人 101 回 / 他人 10100 回である。結果を図 13に示す。

図 13から、単純に平均をとったものよりも FAR が下回っている。これは、入力部分空間を求めることにより、第2主成分以降の情報を用いて類似度(角度)を求めるため、第2主成分方向のデータの分散が考慮され、情報量が増えたために、効果があるものと考えられる。

6 おわりに

本稿では、動画画像を用いた顔認識方法についての報告を行なった。従来、静止画像に対しては部分空間法の有効性が示されており、データの変動を吸収するためには、大量の登録パターンを用いてデータベースを構成することで対応していた。本研究では、認識過程においても、多数次の多様なデータ(時系列画像)を収集し、相互部分空間法を用いて顔認識を行ない、その有効性を確認した。

図1の状況を想定した識別実験においても、パターンの抽出が正確に行なわれていること、さらに環境変動、時間経過が少ない状況を想定したデータであることから、高い識別正解率が得られた。同定実験に関しては、相互部分空間法を用いた場合、他の方法と比較して、評価結果が向上しており、表情変化や顔の向きの影響を吸収していると考えられ、変動にロバストな認識方法であるといえる。

今後は、さらに多人数の場合や、登録からの時間経過と認識率の関連などについて評価を行っていく予定である。

参考文献

- [1] 福井 和宏、山口 修: "形状情報とパターン照合の組合せによる顔特徴点抽出", 信学論 (D-II), vol. J80-D-II, No. 8 掲載予定 (1997)

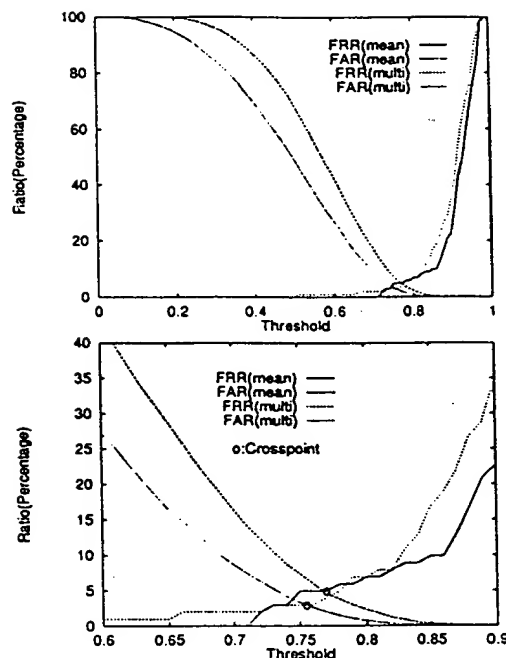


図 13: 平均ベクトルを用いた部分空間法との比較 (下図は拡大図)

- [2] 前田 賢一、渡辺 貞一: "局所的構造を導入したパターン・マッチング法", 信学論 (D), vol. J68-D, No. 3, pp345-352 (1985)
- [3] 宋 炳卓、小沢 慎治: "時系列顔画像処理による個人の認識", 信学技報, PRU93-66, pp. 29-36 (1993)
- [4] 伊藤 典男、富永 英義: "表情の時間的変化を用いた個人識別手法", 信学論 (D-II), vol. J78-D-II, No. 9, pp1315-1324 (1995)
- [5] 石川 則之、有木 康雄: "部分空間法による顔認識の比較実験", 画像の認識・理解シンポジウム MIRU'96, II, 145 (1996)
- [6] E. Oja: "Subspace Methods of Pattern Recognition", Research Studies Press, England (1983)
- [7] K. Fukui: "Edge Extraction Method Based on Separability of Image Features", IEICE Trans. Inf. & Syst., Vol. E78-D, No. 12, pp1533-1538 (1995)
- [8] K. Sung and T. Poggio: "Example-based Learning for View-based Human Face Detection", Technical Report A.I. Memo No. 1521, A.I. Lab., MIT (1994)

- [9] Alex Pentland, Baback Moghaddam, Thad Starner.
"View-based and modular eigenspaces for face recognition". CVPR '94, pp.84-91(1994)
- [10] 佐々木 努、赤松 茂、末永 康仁: "顔画像認識のための色情報を用いた顔の位置合わせ法", 信学技報, IE91-2, pp.9-15(1991)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.